UC Davis

UC Davis Previously Published Works

Title

Quality to impact, text to metadata: Publication and evaluation in the age of metrics

Permalink

https://escholarship.org/uc/item/1pm2s9pg

Journal

Know, 2(2)

ISSN

2473-599X

Author

Biagioli, M

Publication Date

2018-09-01

DOI

10.1086/699152

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Quality to Impact, Text to Metadata: Publication and Evaluation in the Age of Metrics

MARIO BIAGIOLI, UNIVERSITY OF CALIFORNIA, DAVIS

The evaluation of intellectual and scholarly works used to be interpretively complex but technologically simple. One read and evaluated an author's publication, manuscript, or grant proposal together with the evidence it contained or referred to. Scholars have been doing this for centuries, by themselves, from their desks, best if in the proximity of a good library. Peer review—the epitome of academic judgment and its independence—slowly grew from this model of scholarly evaluation by scholars.[1] q1

Things have dramatically changed in recent years. The assessment of scholars and their work may now start and end with a simple Google Scholar search or other quantitative, auditing-like techniques that make reading publications superfluous. This is a world of evaluation not populated by scholars practicing peer review, but by a variety of methods and actors dispersed across academic institutions, data analytics companies, and media outlets tracking anything from citation counts (of books, journals, and conference abstracts) and journal impact factors, to a variety of indicators like H-index, Eigenfactor, CiteScore, SCImago Journal Rank, as well as altmetrics.[2] In most

© 2018 the university of chicago. all rights reserved.

2018009.proof.3d 1 Achorn International 07/03/18 11:12

cases, the sources for these evaluations do not sit on a library's open shelves but in proprietary databases like Clarivate's Web of Science™ or Elsevier's Scopus®.

The roots of this global trend can be traced to the bibliometric techniques initiated by Eugene Garfield in the 1950s. Initially marketed as tools to supplement bibliographic indexes of particular scientific subjects, they were in fact conceived as methods for mapping the structure of scientific and scholarly communities by tracing the networks of citations linking publications within fields and especially across them. Citations were "considered from the point of view of the transmission of ideas"—proxies of intellectual kinship, empirical traces of what scholars read and acknowledged to be relevant to their work, even in faraway fields beyond the reach of specialized bibliographic indexes.[3] Retracing these links (which Garfield referred to as "association-of-ideas indexes" or "thought indexes") would bring to light the intellectual structure of science in its full form—a structure bound to remain hidden from those who instead looked at scientists solely as members of local brick-and-mortar communities like laboratories, departments, or universities.[4] Citation analysis was meant as a sort of X-ray image of the scientific community to visualize otherwise undetectable intellectual networks—the so-called invisible colleges and, later on, Kuhnian paradigms.[5] Garfield went so far as to liken these citation maps to a nonfictional version of H. G. Wells's sci-fi "World Brain"—an all-encompassing information center.[6]

Citation analysis was primarily meant for the scientists themselves to trace ideas throughout the published literature, but this quickly changed in subsequent decades as bibliometrics morphed into a fast-growing and increasingly popular set of tools to evaluate academic quality.[7] By 2010, a *Nature* article could claim that it had "become all but impossible even to count today's metrics."[8] The shift from descriptive to evaluative use of bibliometrics and the vast expansion

2018009.proof.3d 2 Achorn International 07/03/18 11:12

of their range reflected the emergence of new categories of users. Garfield's seminal 1955 paper did not mention deans, which is rather striking given that university administrators across the globe have since become key users of metrics, together with policy makers, politicians, potential donors, students deciding which university they should apply to, and parents trying to put tuition money to best use. We have moved from descriptive metrics used by scientists and scholars, to evaluative metrics used by outsiders who typically do not have technical knowledge of the field they seek to evaluate. This is a shift that reflects a fundamental and increasingly naturalized assumption that the number or frequency of citations received by a publication is, somehow, an index of its quality or value. If early bibliometricians saw citations primarily as links in a field's intellectual *map*, today's metrics focus primarily on citation *counts*.

It is not uncommon to find "value," "importance," and "quality" used interchangeably in contemporary metrics discourse, which implies a rather drastic redefinition of what is meant by these terms. As no reading-based interpretation is involved in bibliometrics, its object cannot be the assessment of scholarly quality as performed by peer review, but some other "value" tied to citations—a value that can be added up in ways that quality cannot. Peer-review evaluation of a scholar's publication is as *singular* as its object, but quantitative evaluations of individual publications based on citations can be aggregated into a score of a scholar's work over a certain review period, which can then be further aggregated with the scores of his/her colleagues to produce a score for the department, and so on all the way up to the global ranking of his/her university. Citation counts and distributions can also be used to assess the prominence of journals and, in the aggregate, of their publishers. No matter the scale or the specific element of the publication system, citations have become the de facto currency of academic value. And, like all currencies, they facilitate transactions, not just evaluations.

2018009.proof.3d 3 Achorn International 07/03/18 11:12

What Has Happened to Evaluation?

Early bibliometrics did not emerge to replace reading but to aid it with more sophisticated search tools. Because the data gathering and analysis was very labor-intensive, the computerization of these tools started early and progressed quickly.[9] It is quite possible that bibliometrics would have never come into being without computers, making it an inherently automated rather than simply quantitative form of knowledge. Greatly facilitated by the advent of electronic publishing, today's metrics-based forms of evaluation depend on evidence gathered by software that scrapes publications' references and metadata such as title, author's name and institutional affiliation, publication date, journal title, and so on. With the exponential growth of computing power and databases, metrics have morphed into a form of "big data" analysis based on algorithms that in some cases rival the complexity of those behind a Google search.[10] To the data analysts who count them, citations have become "community generated content."[11]

Unlike traditional practices of evaluation that, like peer review, are not just qualitative but *craft*-based, metrics cannot be produced by a single scholar but are instead obtained, typically for a fee, from large data analytics corporations—yet another example of today's monetization of data.[12] The introduction of quantitative and automated methodologies has thus introduced a new separation between the producer and the user of the evidence on which the evaluation rests—two roles that were traditionally folded into the same person: the scholar who read and judged. Metrics are therefore a "doubly alien" form of knowledge: both produced and used by people who are not practitioners of the field to which the publications belong.

Metrics have not ushered in a post-truth age of evaluation, but they certainly exemplify a post–peer review regime. It is not that metrics offer no demarcation between good or valuable and bad or useless publications, but rather that they draw that line according to param-

2018009.proof.3d 4 Achorn International 07/03/18 11:12

eters that are not reducible to the interpretation of statements contained in those publications. Citation analysis, for instance, always assigns a *positive* value to citations as events (the fact that somebody did cite something) regardless of whether those citations represent a positive or negative judgment of those publications.[13] That bibliometrics are inherently not about interpreting is demonstrated by the quantitative imagery used by both Garfield (who spoke about a "molecular unit of thought" as the object of a citation)[14] and early fellow bibliometrician Derek de Solla Price (who referred to the content of an article as a "quantum of useful scientific information").[15] Whether they concern number of publications, citations, or "molecular units of thought," metrics are elements of a coherent cosmology of scholarship that conceives both its objects and its methods in quantitative terms.[16] Metrics have thus not *reduced* reading to counting or scholarship to information or data, but have always already conceived them quantitatively, perhaps in the same way that librarians have been naturally inclined to think of knowledge in terms of number of manuscripts, books, or articles held in their libraries, and to conceive of reading as quantified by how many times a book has been checked out.

All proper cosmologies come with their epistemological frameworks. Metric-based evaluations are distinctive in that they are not framed by dichotomies like truth/falsehood or field-specific distinctions like solid/flimsy, original/dated, sophisticated/pedestrian, or elegant/clunky. That such distinctions are contestable and bound to change with the fields that deploy them does not undermine the fact that they are mobilized, at any given point in time, as standards of reference. They are lines that can be drawn differently in different places at different times, but they are still thought of as lines of demarcation between "good" and "bad" publications. Metrics, instead, concern performance—the effects of an action in time. They concern whether a publication (or an author, a department, a university) is impactful

2018009.proof.3d 5 Achorn International 07/03/18 11:12

or not, whether they have made themselves visible or not, and so on. *Metrics are not lines of demarcation, but aggregates of effects*. The increasingly common references to "relevance" in relation to a publication or project share the same logic: relevance is not a form of quality but an indicator of a work's potential for creating effects, that is, of its potential to become impactful.

This is a strangely Cartesian cosmology of scholarship: one in which particles of scholarship move in and out of fields, colliding with other "molecules of thought," producing impact and citations, deviating in new directions toward new collisions, and yet always remaining unitary and unchanged. The faster or bigger the particle, the more impacts it is likely to have. Also in a true Cartesian fashion, this whole scholarly cosmos is not driven by attraction or some other action at a distance, but by the publication's own "quantity of motion," which is imparted to it by the "force" of the journal in which it was published.

From Text to Time: Turning the Publication Inside-Out

Metrics-based evaluation does not simply mark a shift from qualitative to quantitative or, in Aristotelean parlance, from essential qualities to accidents. More radically, it involves a shift to an inherently *time-based* evaluative framework. The traditional evaluation of a scholar's work (say, the folder of a colleague being considered for tenure) involves reading and assessing the content of those texts within the time frame set by the schedule of the evaluation. While that reading has to take place *at some time*, time is effectively external to the object of the evaluation. Publications are treated as objects fixed in print—objects that need to be evaluated "in and of themselves." Instead, unable to read the text, metrics try to trace the *effects* that those publications have had *in time*. Metrics literally step out of the text and into time—the domain of impact. No matter what kind of impact one privileges (ci-

2018009.proof.3d 6 Achorn International 07/03/18 11:12

tations in other publications, citations in patents, journals' impact factors, etc.), the epistemic regime of metrics can only construe the value of a publication as an index of some species of the genus impact.

The shift of evaluative focus from text to time, or from quality to impact, is directly reflected in what counts as relevant evidence. Peer review (which I use as the stand-in for the various forms of qualitative, reading-based, "craft" forms of evaluation) focuses on the "internal" feature of the text, like argument and evidence (sometimes without even considering the author's identity, as in a double-blind review of manuscripts). Metrics, instead, focus on those features of the publication that are external to its claims: title (without which no accounting can begin); authors' names (best if with ORCID digital identifiers, to avoid ambiguities); references (or outgoing citations); journal title (crucial for the "impact factor"); publication date (needed to calculate any type of impact); and institutional affiliations (to help disambiguate authors names and calculate the impactfulness of their universities). q3

Field-changing discussions have taken place in literary studies about the move from modes of reading that focus on the "depth" of a text (starting with the author's intentions and continuing with other deep, hidden, or repressed meanings), against others that instead focus on the "surface" of the text, without assuming the existence of something behind it. One looks at the surface, not through it.[17] The shift from peer review to metrics is arguably more radical as it does not simply involve a shift from depth to surface, but rather moves away from reading altogether, replacing it with harvesting and data mining. It also focuses on elements of the publication that are not just semiotically on the surface, but on those that are literally at its *physical margins*: publication title and date, journal titles, citations, references, authors' names—everything but the content. What is harvested are only the features of the publication necessary to track its impact in time—its metadata.

2018009.proof.3d 7 Achorn International 07/03/18 11:12

A new concept of publication is thus emerging from metrics-based modes of evaluation. Until now, the metadata have been subordinated to the publication, functioning as its frame or as an interface between the publication, its readers, and its catalogers. In the age of metrics, instead, the publication has been recast as a "hook" on which to hang the metadata. It has been literally flipped inside out.

Impact Ontology

Today's multiform varieties of impact were not on the bibliometric horizon when Garfield first introduced citation analysis. The concept of impact, however, was already fundamental to that vision. Properly given citations[18] "are the formal, explicit linkages between papers that have particular points in common."[19] That linkage was effectively conceived as the sign of a publication's "striking" the mind of the scientists who read it, and "bouncing back" in the form of a reference. Like a radar return signal, the citation was not a *representation* of impact but its *material, haptic trace*. It was not a measurement of impact, but was constituted by impact. Impact was the action and the citation was the reaction—two phases of the same process. Even better, a citation was the acknowledged reaction by the reader who was impacted.

This may help to explain why citations are the "gold standard" in today's metrics and rankings. Recent ratings like those produced by *U.S. News & World Report* use the size of a library as a factor in its ranking of law schools, assuming that the number of books on the shelves has a positive impact on the law students who look at them. That, however, is not based on accessible evidence, such as students' feedback or statistical correlations between library size and the students' bar exam pass rate.[20] Taking the raw number of a scholar's publications as an index of impact has comparable limitations, which led an early bibliometrician to dismiss this approach as "counting non-

2018009.proof.3d 8 Achorn International 07/03/18 11:12

sense."[21] Unlike citations that indicate that a publication had an effect, raw publication counts carry no sign of impact, putting them in the same evidentiary category as large law school libraries. An article's download count comes a bit closer to qualifying as an indicator of impact, but does not quite get there either. Because a download implies the agency of someone who clicked on the "download" icon, it can index that action—the downloading—but not the impactful reading of that publication. Compared to these and other indicators of presumed impact, the citation of a publication carries considerably stronger evidentiary value because it functions like a receipt—a statement of impact, if not a proof of effective impact. Still, while the citation indicates that a publication has been engaged, it does not consider the quality of the response—the kind of evidence on which peer review and qualitative evaluation rest.

But if impact has been part of the conceptual apparatus of metrics since their inception, the meaning of impact (and of citation) has changed dramatically. Impact was initially seen as a specific, singular fact: one publication "touched" the mind of one author, causing the production of a material mark—the citation. Ideally, a complete citation index could chart all such links or "thought indexes," producing a map of the distributed collective thinking of a community. Today's metrics, however, ignore the map to simply count the links. No longer a tessera of a thought mosaic whose significance depended on its specificity—a map that looked the way it did because of the specific citations that "drew" it—the citation has become something that is added up and has meaning only in aggregate form. And while originally conceived as a thought index, the citation is now an indicator of the value of impact—an impact that has become a *good* that is *measured* by the citation.

There is a substantial glitch, however. Citations do not measure impact the way bushels measure grain, and barrels measure oil.

2018009.proof.3d 9 Achorn International 07/03/18 11:12

Impact-as-value is treated like a valuable good, but it is not clear what kind of good it is. That impact is also interchangeably referred to as "importance," "significance," "visibility," and sometimes even "quality" points to an unavoidable semantic drift, suggesting that we are no longer dealing with the material cause of a specific citation but with a concept in search of an elusive referent. Citations have thus become reified as numerical icons of an unspecific, nonpresent value. Because value is left undefined, the citations have moved from being the units of measurement of value to becoming valuable tokens in and of themselves; that is, the citation has become the value. We could figuratively think of the citation as a certain amount of "impact ore." No longer a mark of a relation between two publications by means of a concept or thought they share, the citation becomes something that contains value *inside* itself, a value that may be extracted from it through evermore sophisticated data analysis.

This means that while impact used to be the material cause of the citation, *the citation has now become impact.* (By the same token, the citation has come to represent a unit of value of the publication for its author, rather than an index of its effect on the world.) I do not suggest that the material action-and-reaction process linking impact to citation has started to run in reverse, but that a very different discursive framework has developed since early bibliometrics, one that treats impact as value and turns citations into a representation of such a value. Nobody, however, has been able to conceptualize—let alone demonstrate—the specific connection between impact measured through citations and the value or quality of a publication.[22] While such a link is widely assumed to exist (the justification of assessing scholarly work through quantitative indicators as proxies of quality depends on that), the fact that it still remains unknown despite the fundamental role attributed to it suggests that the presumed connection between impact and "quality" is a discursive placebo. The metrics episteme is only about effects. It is citations all the way down.

2018009.proof.3d 10 Achorn International 07/03/18 11:12

Playing with the Timeline

While impact is inherently historical (conceptually defined as an effect and empirically assessed on the basis of past performance), it is in fact expected to function as a predictor of future performance—the kind of evidence one uses to determine an investment's potential risks and rewards. Past impact becomes a forward-looking statement.

This irresolvable tension is most clear in the journal impact factor (JIF), possibly the most influential indicator of impact today. Publications are deemed to gain impact or value when they are cited copiously and often, but also when they are published in high-quality journals. A high-quality journal is visible and widely read, which lends visibility to the articles it publishes, making them more likely to be read and cited. An article also gains value from being published in a high-quality journal because that is taken to mean that it survived a particularly demanding peer review process, which the journal needs to adopt due to the very high number of submissions it attracts because of its quality and visibility. A journal's impact factor tracks how many citations all the articles published in that journal have received in a two-year period, divided by the number of citable articles it published in that same period. It is an indicator of the average "density" of citations received by its articles.

Publication used to be clearly separate from evaluation—there needs to be a publication before one can evaluate it. When it relies on the journal's impact factor, however, evaluation no longer follows publication but becomes paradoxically simultaneous to it. That is because this type of evaluation depends on identifying the venue of the publication and attaching an index of that location—the impact factor—to the publication. A publication is born evaluated, making the JIF look like a strange aristocratic title bestowed at birth (based on the name of the journal where the publication was born) rather than gained "meritocratically" during the publication's life.

2018009.proof.3d 11 Achorn International 07/03/18 11:12

This subverts the very notion of impact. Impact refers to an effect, that is, to something that has already happened, like the citations an article has received since its publication. The increasingly coveted JIF, however, functions as an estimation of impact *before it happens*, a device for producing an instantaneous evaluation of a publication that can in fact only accrue value (i.e., impact) in the future. This is qualitatively different from saying that the value of things is bound to fluctuate in time, and that the impact factor is an estimation of that fluctuating value. A house has value both when it is first built and years after that, but, by definition, the impact of a publication *does not and cannot exist* when the publication comes off the press. The impact factor, therefore, can neither measure the value of a publication's impact nor estimate its future value based on its present value (the way one may estimate the future value of a house based on its present features). The impact factor does not estimate a publication's value but construes it, and does so not based on the features of that publication but on the citations received by unrelated articles published in that journal over a certain period, in the past.

The acrobatic manipulation of the timeline involved in using the JIF to evaluate a publication is probably justified by the useful effects (rather than the accuracy) of such an evaluation. It prices the article (and thus the "value" of its authors) right now rather than years down the road, after its citations could mature and be harvested. It is a rather crude tool to price futures, which I do not mean as a metaphor: some universities hand out substantial cash bonuses to their faculty for their publications, indexing the bonuses on the journals' impact factor. In China, *Nature* and *Science* articles fetch, on average, $43,000 a piece.[23] These universities probably justify such bonuses by expecting that the impact factors of their faculty's publications will improve the institutions' future rating. Impact is thus not just a valuable good, but one that can be transacted: I can buy my faculty's impact and then sell it back (in the form of tuition fees) to more students who will enroll at

2018009.proof.3d 12 Achorn International 07/03/18 11:12

my university because of its higher ratings. It is by playing with the timeline that the JIF helps to sustain a faster pace of transactions in a global market of academic value based on impact. It produces impact that has not happened yet, value that has no value yet, showing that metrics do not assess but rather create value. Impact has become *Impact*™.

Rankings as Competitive Episteme

Metrics are inseparable from rankings—not in the trivial sense that metrics produce rankings, but that, counterintuitively, rankings give meaning to metrics rather than the other way around. Scholars have been ranking other scholars (dead or alive, formally or informally) for a long time. If I serve on a search committee for a senior faculty appointment, I will be asked to read the top scholars in my field and rank them. Obviously, reading and evaluating precede ranking. But not so in metrics concerning scholarly publications, where ranking comes first and provides the condition of possibility for evaluation. Considered by itself, the fact that one of my articles has received 100 citations is just a number, but it becomes meaningful as soon as that number is compared to the 200 citations received by my colleague's article. The fact that I have fewer citations than my colleague creates the assumption that my publication has been the less impactful no matter how accurate or inaccurate the translation between citation counts and value may be. It is the fact that I end up ranked second that turns the number 100 into something closer to an evaluation of my publication, even though no clear "exchange rate" between citations and impact is ever given. A ranked list casts the quantitative differences between entries as meaningful.

Quantity matters when it comes to comparisons. Longer lists of rankings produce stronger reality effects because they intimate that a large population has been canvassed, which in turn suggests that

2018009.proof.3d 13 Achorn International 07/03/18 11:12

what is being ranked is a quality that exists throughout that population—something not likely to be accidental or meaningless. Shifting focus from one publication to a comparison between two or more of them in no way settles the question of how citations are connected to value, but it does help to avoid the deadlock that would be created by insisting on questioning the specific nature of that connection. That impact and value are left undefined does not stop one from cranking out rankings anyway. Producing one ranking after another allows one to continue to assume that there is a relation between quantity of citations and value or quality without ever showing what that relation is. It is as if the virtual completeness of the sample substitutes for the absence of a conceptual unity in the ranking.

This incentivizes the production of more comprehensive rankings, but also the ranking of different kinds of things—publications, journals, departments, universities, and so on. While all these rankings of different things are technically ungrounded—there is no clear correspondence between rankings and quality—they support each other by creating the impression that we inhabit a cosmos of rankings, where everything is rankable and rankings naturally index some kind of value. To a large extent, this proliferation is market-driven. There are, for instance, competing rankings of global universities that emphasize different features that benefit or penalize certain countries and institutional profiles. But the fact that specific rankings compete in the global ratings marketplace is distinct from (if overlapping with) the fact that the ranking episteme is one that both incentivizes and is made more credible by the existence of more rankings of more things.

Competition among ranking agencies does not just incentivize the production of more and more comprehensive rankings, but structures their very logic. The assumption of a relation between citation counts and impact is inherently dependent on comparisons, which means that impactfulness is always relative and competitively constructed. There is therefore something distinctly recursive about the ranking

2018009.proof.3d 14 Achorn International 07/03/18 11:12

episteme: the relation between citations, impact, and "quality" is constantly *deferred* through the very repetition and expansion of comparisons and rankings. That relation is neither refuted nor confirmed, but simply displaced into the future.

The comparative and generative nature of rankings creates a pressure to produce not just more citation-based rankings but more citations themselves. This has clear inflationary effects. Consider a fictional scenario in which, a few years ago, my university decided that in order to maintain our excellent position in the global rankings, each member of the faculty should aim at producing publications that yield fifty citations a year. It is easy to imagine how a lower-ranking university that wants to overtake us in the global rankings would tell its faculty that they need to produce publications yielding seventy citations a year. And as soon as our competitors start to overtake us in the rankings, my university would of course respond by asking its faculty to aim for eighty citations a year, presumably by publishing only in journals with the highest impact factor. Precisely because citation counts cannot be reliably translated into "quality," the only variable that can be maximized is the citation count itself, no matter what that count really signifies. The meaning of rankings is created through competitive comparison, rankings increase competition, and competition pushes the quantitative standards of the competition higher and higher, with no external term of reference to establish what "too much" or "too little" may mean. It is competitive comparison all the way down, and up.

Metrics of Excellence

The increasing reliance on citation-based indicators is connected to the omnipresence of "excellence" in academic policy discourse. The university has grown into an institution that does too many things with and for too many constituencies to remain identifiable with the simpler and clearly unified educational mission of the nineteenth-century

2018009.proof.3d 15 Achorn International 07/03/18 11:12

Humboldtian university, which provided the template for the then-emerging research-based institutions of higher education in the United States. Bill Readings has convincingly argued that we no longer have either a unified "idea" of the university or a shared definition of a holistic academic education.[24] This profound identity crisis also affects the definition of academic quality, which has become literally unmoored. Excellence has emerged as a response to this crisis, embodying a redefined notion of "quality" for the post-Humboldtian, modern university. To achieve excellence does not mean to achieve quality according to an external, stable term of reference, but simply to be great at whatever one is doing. A university can achieve excellence in philosophy as well as in parking services.[25]

Excellence sounds like quality but is inherently about performance, that is, about impact. The replacement of quality with excellence in discussions about the value of the modern university is thus logically equivalent to replacing "qualitative quality" with unspecific notions of impact in the evaluation of scholarly work. Discourses both of academic excellence and of impact emerge from the demise of a unified idea of quality. And in both cases the shift toward excellence or impact hinges on and spawns the production of rankings. A university is no longer "good" or "bad" in a general sense, but is ranked (rather than evaluated) in comparison to other universities. And the term of comparison is not one unified notion of quality but an indefinitely long list of quantifiable features: most Nobel Prize winners, highest publication output, most grants, highest level of employment among recent graduates, most diverse student population, most cited faculty, largest library, best-lit stadiums, fattest squirrels,[26] and so on.

Far from being a problem, the fact that excellence has no inherent referent becomes a powerful discursive tool when it can be framed and operationalized through metrics and rankings. Quantitative comparative techniques give a sense or effect of specificity to excellence by tying it to specific types of impact or results without, however,

2018009.proof.3d 16 Achorn International 07/03/18 11:12

showing (or even attempting to show) how those specific forms of excellence or impact add up to quality. Conversely, when viewed from within the discourse of excellence, the proliferation of different rankings does not signal a lack of analytic focus but rather a demonstration that excellence comes in many different shapes and colors.

Ranking for Investment

Why has the peculiar kind of evaluation provided by metrics—an estimate of so-called impact—become so popular? To whom does that impact matter if it does not seem to be relevant to the scholars themselves? Why and where is impact-based evaluation impactful? The short answer is that the type of evaluation provided by metrics has an uncanny fit with decisions about resource allocation. These decisions are distinctly distributive in nature: less about "yes" or "no" and more about "how little" or "how much"—to whom. For example, introducing the journal impact factor in 1972, Garfield presented it, among other potential uses, as a tool to help librarians in their wise allocation of their subscription budget by picking the most significant journals.[27] The following year saw the publication of the first National Science Foundation (NSF) report on science indicators, which drew a connection between bibliometric evidence and science policy and funding decisions.[28] In subsequent decades, countries ranging from the United Kingdom to Denmark, Italy, and New Zealand have tied government funding of universities to assessments of their research performance, which have become increasingly metrics-based.[29]

The logic behind these programs seems to be that of investment, which is closely related to impact. The neoliberal perspective that animates this trend sees government funding not as a contribution owed to the university according to some traditional "social contract," but as a public investment in the university. Accordingly, these funds are not means for the university's maintenance of its status quo (to

2018009.proof.3d 17 Achorn International 07/03/18 11:12

which it is no longer seen to be entitled), but as resources to do more and better *in the future*. Based on these premises, it would then seem appropriate to allocate funds according to a bundle of metrics of the estimated impact that the university's publications and teaching have had. The key issue is not whether those publications were "high quality" in some general sense, but whether their impact (or, more often, narratives of impact developed on that evidence) indicate that the university will be able to deliver the kind of performance expected from the allocation of those funds. In a context driven by the discourse of innovation, quality may look too outdated a target—too static to capture the evolving goals of the entrepreneurial university.

Metrics about past productivity (number of publications, citations, grants, patents, etc.) are thus mobilized as indicators of the university's potential as a future partner for external investors. This includes the students and their parents who become eager consumers of rankings when deciding where to apply and invest their tuition money. Universities whose quality has been a matter of tradition remain the most attractive, but students may be willing to bet on institutions whose rankings are not yet stellar but ticking upward, like a growth stock. No matter the specific object of academic metrics, their perspective is always trained on future potentials through the lens of present results.

Incommensurabilities of Shared Governance

The use of metrics to evaluate scholarship is routinely described as a form of audit. An audit involves an evaluation by a third party with no personal or financial ties to the audited, or the same knowledge. One can audit NASA without being a rocket scientist, or the finances of a university without being an academic. Articulated from the conceptual template provided by accounting, the auditors' charge is not to assess the products of the audited parties, but the accuracy of the re-

2018009.proof.3d 18 Achorn International 07/03/18 11:12

ports they submit (accounting books, financial reports, etc.) to document their activities. Auditing is not about assessing whether a given car manufacturer produces good or bad cars, but whether it is providing its stakeholders with accurate representations of its performance in relation to its corporate targets. Even when quality is the object of the auditing—as in the so-called Quality Management Systems—what is checked is not the quality of the product, but the systems and procedures a certain company has put in place to ensure quality control.[30]

Like water and oil, the logic of the audit and that of peer review do not mix. Audits concern the review of *representations of operations* (like accounting books) according to standard protocols, but peer review produces evaluations of the *products of operations* (like scholarship) according to local and possibly nonexplicit practices. Further, the audit epitomizes detached knowledge, predicated on a firewall between auditor and audited, while peer review rests on the assumption that the only judgment that matters is that of peers. Scientists evaluate other scientists' claims by replicating their experiments or by going over their data with their same statistical tools. Historians analyze the same documents and sources used by their colleagues in order to assess their claims, and mathematicians check other mathematicians' proofs by redoing them. As there is virtually no distinction between the practices of the auditors and the audited, between the practices scholars employ to do their work and those they use to evaluate the work of their colleagues, peer review epitomizes insiders' knowledge. Accordingly, there can be no other relevant knowledge except that of peers, which implies that there can be no conditions of possibility or conceptual space for the audit.

There is no need to uncritically accept the representation of peer review as the best way to evaluate scholarly work, or the fairest tool to distribute recognition among scholars. Peer review has been shown to reward conformity over innovation, enable power politics through

2018009.proof.3d 19 Achorn International 07/03/18 11:12

peer pressure, and facilitate discrimination and even plagiarism under the cloak of anonymous reports. It is also known to fail both to spot fraud and to properly judge quality (as shown by the rejection of articles subsequently cited in Nobel Prize awards). Peer review is probably just the least bad mechanism of evaluation we have developed so far. The point, therefore, is not to defend it from metrics' attempted takeover, but to recognize that we are looking at two competing regimes of evaluation that are virtually incommensurable. One casts evaluation as a necessarily external audit while the other construes it as a necessarily internal judgment. One looks at the inside or content of a publication, while the other looks at its margins, or metadata.

The incommensurability between these two regimes was largely unproblematic in the past when the use of metrics was limited to decisions about high-level resource allocation, based on data of a granularity that was far too coarse to even think about using it to evaluate individual authors and publications. The principle of shared governance that frames most US research universities allowed the introduction of metrics as a form of knowledge that was not above peer review but complementary to it. Peer review judged scholarly matters, while metrics informed the administrators' budgetary decisions. There was never a "balance" between peer review and metrics (more like a separation between church and state), but this sharp demarcation started to blur, and create conflicts, following the progressive expansion of the use of metrics to evaluate individual authors and publication.

But how did metrics manage to take over some of the domain of peer review if the former cannot demonstrate any specific epistemic superiority over the latter? Metrics seem to be winning not because they have proven their superiority over peer review, but because they have successfully developed many more users. Their weapons are not demonstrably better, but they are surely easier to use, thus enabling a

2018009.proof.3d 20 Achorn International 07/03/18 11:12

much larger army. Metrics-based evaluations of scholarly work may be used with little or no knowledge of the publications or the disciplines they belong to, and may be easily accessed online, with the appropriate subscription. Instead, due to its local and skill-intensive practices, peer review is bound to keep the number of qualified evaluators down to the members of that specific field and academic craft. Scenarios are thus emerging in which *many "alien" judges* armed with quantitative techniques, computer-generated statistics, and the objectivity derived from distance and lack of investment in the field and its objects confront *many fewer "insiders"* who rely on qualitative forms of judgment whose parameters are opaque to all but those insiders—a judgment they apply to objects they are close to and passionate about.

Moral Numbers

Unlike the incommensurability that renders debate ineffectual by having the participants talk past each other, the radical epistemic difference between scholar-made qualitative evaluation and software-calculated metric indicators have specific institutional effects within academic shared governance: metrics allow actors external to a field of expertise to make decisions that go against the opinions and proposals of the experts, but do so without directly challenging their knowledge. For example, librarians can use impact factors to decide which journals are worth subscribing to, and do so without any knowledge of any of the articles published in those journals and without the need to argue against the humanities faculty who complain about the cancellation of the subscription to journals they deem excellent. The librarians can simply say, without in any way questioning the faculty's assessment of those journals, that they still do not have a sufficiently high impact factor (and thus potential readers) to justify the expense.

2018009.proof.3d 21 Achorn International 07/03/18 11:12

When deployed within a shared governance framework, the use of metrics introduces an important asymmetry between two methodologies that are otherwise not rankable due to their incommensurability—an asymmetry that then becomes value-laden. For instance, metrics contribute to making the decisions of the administrators (or more generally of those taking the auditor's role) appear objective and impersonal by casting them as based on transparent calculations and thus distinct from the argument-based (or even "argumentative") judgments of scholars practicing peer review. Metrics and peer review are presented as complementary, but when consensus is not achieved, metrics discourse manages to come across as objective and moral, even though it cannot be shown to be grounded in better and sounder evidence and methodology.

Theodore Porter has shown that people are likely to trust numbers when they do not trust one another.[31] The increasing reliance on metrics of academic performance fits Porter's general claim, with a twist: metrics help to bypass the problem of personal mistrust, but do so by rechanneling that distrust in the direction of those who do not use metrics like, in this case, the users and supporters of peer review. Not only is peer review "argumentative" and "opaque" rather than "transparent," but its practitioners—because of the requirement that one has to be a peer in order to judge—are liable to be represented as a closed community with high costs of entry and little interest in being accountable to outsiders and, by extension, to the broader public and the taxpayer.[32] Compared to the proponents of metrics, the supporters of peer review can be made to look like they are defending their privileged knowledge, trying to prevent it from becoming the object of an audit.

Peer review epitomizes specialist knowledge, which is the knowledge of a small community. Depending on where you stand, it can be represented as the most sophisticated knowledge one can develop about that topic, or as a source of bias, like Bacon's "idols of the cave."

2018009.proof.3d 22 Achorn International 07/03/18 11:12

To audiences unwilling or unable to appreciate that communities of expertise tend to be small and relatively closed because of the remarkably lengthy training required to gain membership in them, academic departments may look like incestuous tribes, dens of special interest and myopic self-referentiality, with a tendency both to think in a certain way and to hold onto specific local interests. Metrics promise to bypass all of that, sorting true knowledge from mere statements of privilege dressed up as local expert knowledge. Metrics are knowledge by and for outsiders, understandable to all those who consider themselves external stakeholders in academic scholarship, all the way down to the taxpayers. Metrics can even acknowledge some of their limitations because their transparent methodology allows for criticism, and therefore for corrections. Though not epistemologically superior, metrics claim the rhetorical moral upper hand.

But where does this specific notion of morality come from? How did accountability become the mantra of academic administrative discourse? The demand for accountability sounds routine today, but it is a recent development connected to large-scale resource allocation involving government funds to all universities (in the United Kingdom) or state funding to state universities. In that context, accountability had a clear reference: the taxpayers. The taxpayers provided the funds that government and state officials distributed to the universities, and the taxpayers deserved to know what was done with their money.

Invocations of accountability, however, are now found everywhere, well beyond scenarios involving scholarship supported by public funds. Along the way, the discourse of accountability has largely shed specific references to the persons or entities that scholarship and scholars ought to be accountable to. In other words, accountability has developed a bewildering range of meanings while shedding specific referents, making it ultimately look as generic as excellence.[33] But, as with excellence, the semantic drift does not devalue the perceived

2018009.proof.3d 23 Achorn International 07/03/18 11:12

importance of accountability. On the contrary, it makes it more valuable by expanding its potential applicability. It has become good to be accountable, no matter who or what one should be accountable for and to. Generic or potential accountability has thus morphed into a virtuous (and virtual) state: "being accountable."

Unmoored from its referent, "being accountable" sounds like a moral value, though in this context, "value" is more closely connected to "valuable" than to "moral." Being accountable no longer refers to a specific relation of responsibility between specific knowledge producers or knowledge claims and the specific people who use that knowledge. Rather, it functions like a badge or mark of certification indicating the additional value that a certain methodology like metrics has by virtue of having the capacity for accountability—a capacity it may or may not actualize. Without being demonstrably more accountable than peer review, metrics are succeeding at claiming the *Accountability*™ brand for itself.

2018009.proof.3d 24 Achorn International 07/03/18 11:12

Notes

I wish to thank Davide Proserpio, Kriss Ravetto-Biagioli, and the participants at the SIFK Inaugural Conference on "Practices of Knowledge" for all the comments and criticism.

1. Mario Biagioli, "From Book Censorship to Academic Peer Review," *Emergences* 12 (2002): 11–45; Alex Csizar, *The Scientific Journal: Authorship and the Politics of Knowledge in the Nineteenth Century* (Chicago: University of Chicago Press, 2018).

2. Michael Power, *Audit Society: Rituals of Verification* (Oxford: Oxford University Press, 1997); Marilyn Strathern, ed., *Audit Cultures: Anthropological Studies in Accountability, Ethics and the Academy* (London: Routledge, 2000).

3. Eugene Garfield, "Citation Indexes for Science: A New Dimension in Documentation through Association of Ideas," *Science* 122 (1955): 109–10.

4. Garfield, "Citation Indexes for Science," 108.

5. Derek De Solla Price, *Little Science, Big Science* (New York: Columbia University Press, 1963), 62–91; Eugene Garfield, Morton V. Malin, and Henry Small, "Citation Data as Science Indicators," in *Toward a Metric of Science: Advent of Science Indicators*, ed. Yehuda Elkana, Joshua Lederberg, Robert Merton, Arnold Thackray, and Harriet Zuckerman (New York: Wiley, 1978), 183, 193.

6. Eugene Garfield, "'Science Citation Index'—a New Dimension in Indexing," *Science* 144, no. 3619 (May 8, 1964): 649–54, at 649.

7. He thought that citation counts would provide key evidence for historians interested in studying the influence of past scientists and publications—the "thinking of the period"—but such evaluations were scholarly, not professional (Garfield, "Citation Indexes for Science," 110).

8. Richard Van Noorden, "A Proliferation of Measures," *Nature* 465 (2010): 864. For a recent overview, see Roberto Todeschini and Alberto Baccini, *Handbook of Bibliometric Indicators: Quantitative Tools for Studying and Evaluating Research* (Weinheim: Wiley-VCH, 2016).

2018009.proof.3d 25 Achorn International 07/03/18 11:12

9. Garfield, "Citation Indexes for Science," 109; Derek de Solla Price, "Networks of Scientific Papers," *Science* 149 (1965): 510–15.

10. The so-called Eigenfactor is expressly modeled after the Google PageRank model (http://www.eigenfactor.org/about.php). The family of indicators that go under the name of "altmetrics" also involve complex, multifactor indexes.

11. James Pringle, "Trends in the Uses of ISI Citation Databases for Evaluation," *Learned Publishing* 21 (2008): 86. The author is Vice President, Product Development, of Thomson Scientific—the corporate heir to Garfield's Institute of Science Information.

12. Yves Gingras, *Bibliometrics and Research Evaluation: Uses and Abuses* (Cambridge, MA: MIT Press, 2016), 1–10.

13. "Important should not be confused with correct, for an idea need not be correct to be important" (Garfield, Malin, and Small, "Citation Data as Science Indicators," 182).

14. Garfield, "Citation Indexes for Science," 108.

15. De Solla Price, *Little Science, Big Science*, 62.

16. Bibliometrics cannot even determine that the citation is conceptually specific to what it refers to. H. G. Small ("Cited Documents as Concept Symbols," *Social Studies of Science* 8 [1978]: 327–40) addressed this issue, in my view, without success.

17. Stephen Best and Sharon Marcus, "Surface Reading: An Introduction," *Representations* 108 (2009): 1–21.

18. That is, not those citations given to friends and patrons without even reading their work. An excellent analysis of citation ethics is in Roald Hoffmann, Artyom A. Kabanov, Andrey A. Golov, and Davide M. Proserpio, "Homo Citans and Carbon Allotropes: For an Ethics of Citation," *Angewandte Chemie International Edition* 55 (2016): 10962–76.

19. Eugene Garfield, *Citation Indexing: Its Theory and Application in Science, Technology and Humanities* (New York: Wiley, 1979), 1.

20. *U.S. News & World Report*, "Methodology: 2018 Best Law School Rankings," https://www.usnews.com/education/best-graduate-schools/articles/law-schools-methodology (under "Faculty Resources").

21. De Solla Price, *Little Science, Big Science*, 62.

22. For an extensive review of the literature on the topic, see Lutz Bornmann and Hans-Dieter Daniel, "What Do Citation Counts Measure? A Review of Studies on Citing Behavior," *Journal of Documentation* 64 (2008): 45–80.

2018009.proof.3d 26 Achorn International 07/03/18 11:12

23. Alison Abritis et al., "Cash Bonuses for Papers Go Global," *Science* 357, no. 6351 (August 11, 2017): 541.

24. Bill Readings, *The University in Ruins* (Cambridge, MA: Harvard University Press, 1997), 44–69.

25. Readings, *University in Ruins*, 21–43.

26. "The quality of an institution of higher learning can often be determined by the size, health and behavior of the squirrel population on campus," at: http://www.gottshall.com/squirrels/campsq.htm.

27. Eugene Garfield, "Citation Analysis as a Tool in Journal Evaluation: Journals Can Be Ranked by Frequency and Impact of Citations for Science Policy Studies," *Science* 178 (1972): 471–79, at 477.

28. National Science Foundation, *Science Indicators 1972* (Washington, DC: US Government Printing Office, 1973).

29. James Wilsdon, *The Metric Tide* (London: Sage, 2016), 23–28.

30. "A quality management system (QMS) is a set of policies, processes and procedures required for planning and execution (production/development/service) in the core business area of an organization" at: http://the9000store.com/iso-9001-2015-requirements/what-is-iso-9001-quality-management-system/.

31. Theodore M. Porter, *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life* (Princeton, NJ: Princeton University Press, 1995).

32. Cris Shore and Susan Wright, "Coercive Accountability: The Rise of Audit Culture in Higher Education," in *Audit Cultures*, ed. Marilyn Strathern (London: Routledge, 2000), 69.

33. "Public Accountability," in *The Oxford Handbook of Public Accountability*, ed. Mark Bovens, Robert E. Goodin, and Thomas Schillemans (Oxford: Oxford University Press, 2014), 4–7.

2018009.proof.3d 27 Achorn International 07/03/18 11:12

QUERIES TO THE AUTHOR

Q1. Au: Your article has been lightly edited for grammar, clarity, consistency, and conformity to journal style, including issues of hyphenation and capitalization. The *Chicago Manual of Style* is followed for matters of style, and *Merriam-Webster's Dictionary* is followed for spelling. Please read your proof carefully to make sure that your meaning has been retained.
Q2. Au: For consistency with the majority of usage elsewhere, we have treated "metrics" as a plural term throughout. OK?
Q3. Au: Should ORCID be defined at first mention?

2018009.proof.3d 28 Achorn International 07/03/18 11:12